

Localization from superselection rules in translation invariant systems

Isaac H. Kim^{1,2} and Jeongwan Haah³

¹*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada*

²*Institute for Quantum Computing, University of Waterloo, Waterloo ON N2L 3G1, Canada*

³*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
(Dated: May 8, 2015)

We study a translation invariant spin model in a three-dimensional regular lattice, called the cubic code model, in the presence of arbitrary extensive perturbations. Below a critical perturbation strength, we show that most states with finite energy are localized; the overwhelming majority of such states have energy concentrated around a finite number of defects, and remain so for a time that is near-exponential in the distance between the defects. This phenomenon is due to an emergent superselection rule and does not require any disorder. An extensive number of local integrals of motion for these finite energy sectors are identified as well. Our analysis extends more generally to systems with immobile topological excitations.

I. INTRODUCTION

Recently there has been a significant amount of interest on the mechanism behind how isolated quantum systems thermalize. While the eigenstate thermalization hypothesis asserts that quantum many-body systems thermalize at the level of the eigenstates,^{1,2} there are well-known counterexamples to this proposal as well. In particular, a seminal work of Anderson shows that disorder can inhibit thermalization for noninteracting systems, as particles are localized indefinitely at some fixed location.³ Recent studies indicate that the effect of localization can persist in the presence of interaction, even at finite energy densities.⁴⁻⁷ These results suggest that disorder can cause the system to “freeze” in time, and hinders equilibration. There are also some translation invariant systems which were proposed to be localized under a generic disordered initial conditions.⁸⁻¹³ However, currently there does not seem to be a decisive consensus as to whether such effects persist in the thermodynamic limit; see Ref. 14.

In this paper, we point out that interacting quantum many-body systems can be localized by a completely different mechanism. This is due to an emergent superselection rule, and does not rely on disorder. We consider an exactly solvable and translation-invariant lattice model of spins, called the cubic code model¹⁵, and find a set of states with manifestly localized spatial energy profile. These states remain invariant under Hamiltonian evolution, and constitute the overwhelming majority of the states at finite energy. Our main result here is that all these properties remain stable against *arbitrary perturbations*, so long as the perturbation strength is smaller than some finite critical value. Hence we conclude that this phenomenon is a robust property of the phase.

To be more precise, we consider a weak arbitrary perturbation, denoted as Y , to the cubic code model H_0 that shall be defined later, and construct a set of states $|\psi_Y\rangle$ such that their energy is concentrated around a finite number of points with respect to the perturbed Hamiltonian $H_0 + Y$. These states are quasi-eigenstates in the

sense that

$$\left| \langle \psi_Y | e^{-it(H_0+Y)} | \psi_Y \rangle \right| \geq 1 - tL^\alpha \exp(-cL^\eta) \quad (1)$$

where $\alpha, c, \eta > 0$ are constants and L is the smallest distance between the defects. Furthermore, the fraction of these states within the low-energy subspace is shown to approach 1 in the thermodynamic limit.

While we only consider a system with finite number of defects, we emphasize that our system is fundamentally different from noninteracting fermionic systems. The transport of the low-energy excitations are hindered by a novel form of emergent superselection rule, which does not arise in free systems. Interestingly, localization originates from strong interaction within our setup, contrary to what is observed in the context of disordered systems.

It should be noted that the localized states satisfying Eq. (1) are unlikely to exist in a system with non-trivial dispersion relation. Consider a one-dimensional fermionic chain with a Hamiltonian $H = \int dk \epsilon(k) c_k^\dagger c_k$ as an example. Here, c_k is the annihilation operator of momentum k . If the dispersion relation is $\epsilon(k) = -\lambda \cos k$, then a state representing a single localized excitation, such as $|\psi\rangle = \int \frac{dk}{\sqrt{2\pi}} c_k^\dagger |\Omega\rangle$, will spread out quickly in time;

$$\langle \psi | e^{-itH} | \psi \rangle = J_0(\lambda t) = 1 - \frac{\lambda^2 t^2}{4} + \dots,$$

where $J_0(t)$ is the Bessel function.

In the rest of the paper, we shall introduce our model and explain how Eq. (1) is derived. We first begin by explaining the central concept, the *locally gapped state*. Roughly speaking, the notion of locally gapped state formalizes the intuition that two eigenstates do not mix with each other with respect to a perturbation if they are either (i) separated from each other with a large energy gap or (ii) the corresponding off-diagonal matrix element of the perturbation is sufficiently small. We then proceed by identifying all the low-energy states of our model which are locally gapped. A simple estimate shows that locally gapped states form the overwhelming majority of

the states within the low-energy subspace. Lastly, we rigorously prove that the aforementioned properties remain stable in the presence of weak perturbations, and find an extensive set of approximate local integrals of motion. Throughout the paper, we shall adopt the following technical terminology: A function $f(x)$ *rapidly decays* if it scales like $x^\alpha \exp(-cx^\beta)$ for large x , where $c, \beta > 0$ and α are constants.

II. LOCALLY GAPPED STATES

With respect to a Hamiltonian H , a state $|\psi\rangle$ is said to be *locally gapped* with a diameter d and an energy Δ , if for any orthogonal state $|\phi\rangle$, one of the following holds:

- $|\langle\phi|H|\phi\rangle - \langle\psi|H|\psi\rangle| \geq \Delta > 0$
- $\langle\phi|O|\psi\rangle = 0$ whenever O is supported on a ball of diameter d .

This means that $|\psi\rangle$ is separated by an energy gap of Δ from all the other states that are reachable by local operators from $|\psi\rangle$. Unless specified otherwise, we will be interested in locally gapped states with an extensive diameter and finite energy gap. Let us mention some examples. A non-degenerate ground state of a Hamiltonian with an energy gap is locally gapped. A less trivial example is the ground state sector of topologically ordered systems. The ground states can be only connected to each other by an operator whose support is extended across the system.

For clarity, let us also discuss states which are *not* locally gapped. Consider a Hamiltonian $H = \sum_i \sigma_i^z$ on a chain where σ^z is the Pauli- z matrix. It is gapped with a unique ground state in which all the spins are pointing “down.” Its first excited state is a spin configuration with a single spin, say the j th one, pointing up. This excited state is not locally gapped because there is another excited state for which only the $(j+1)$ th spin is pointing up. These two are related to each other by a local operator, namely $\sigma_j^x \sigma_{j+1}^x$. These states are not locally gapped with diameter $d > 1$.

The above conclusion exemplifies a general observation: If the hopping of a quasi-particle is realized by a local operator, states with localized excitations are not locally gapped. This intuition is true for many translation-invariant systems, and it is thus tempting to conjecture that the only way to have a locally gapped excited state is through some strong disorder. However, we shall show that there exists a translation-invariant system with many locally gapped excited states.

One might want to consider an approximate and more general version of locally gapped states by allowing the transition amplitude $\langle\phi|O|\psi\rangle$ to be approximately equal to zero for local operators O . We shall encounter this generalization later, but for now it will only make the notation more cumbersome.

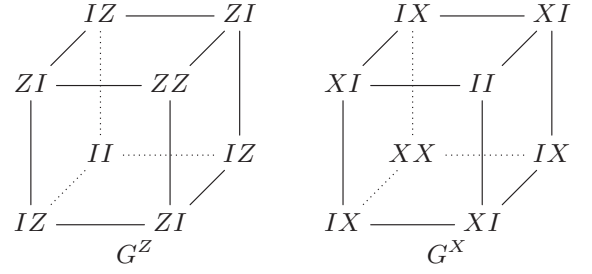


FIG. 1. Cubic code model. X and Z represent Pauli matrices σ^x and σ^z , respectively. A term in the Hamiltonian is the product of 8 Pauli matrices arranged as in the diagram.

III. LOCALLY GAPPED STATES IN THE PURE CUBIC CODE HAMILTONIAN

The cubic code model is defined on a simple cubic lattice with two qubits per site, where the Hamiltonian is a translation-invariant sum of the two types of interaction terms¹⁵.

$$H_0 = -J \sum_i G_i^Z + G_i^X \quad (J > 0) \quad (2)$$

$$G_i^Z = \sigma_{i,1}^z \sigma_{i,2}^z \sigma_{i-\hat{x},1}^z \sigma_{i-\hat{y},1}^z \sigma_{i-\hat{z},1}^z \sigma_{i-\hat{y}-\hat{z},2}^z \sigma_{i-\hat{z}-\hat{x},2}^z \sigma_{i-\hat{x}-\hat{y},2}^z$$

$$G_i^X = \sigma_{i,1}^x \sigma_{i,2}^x \sigma_{i+\hat{x},2}^x \sigma_{i+\hat{y},2}^x \sigma_{i+\hat{z},2}^x \sigma_{i+\hat{y}+\hat{z},1}^x \sigma_{i+\hat{z}+\hat{x},1}^x \sigma_{i+\hat{x}+\hat{y},1}^x$$

See Fig. 1 for the arrangement of the Pauli matrices in the terms. Every term in the Hamiltonian commutes with each other, and as such, the energy spectrum is discrete. For the ground state, both G_i^X and G_i^Z take eigenvalues of $+1$. The ground state subspace is degenerate and is topologically ordered in the sense that any local operator has the same expectation value for all the ground states. According to our definition in the previous Section, the ground state is locally gapped by an energy $\Delta = 2J$ and diameter $L - 1$, where L is the linear system size. The excited states can be described by *defects*, which are the violated local terms, e.g., $G_i^X = -1$ or $G_i^Z = -1$. We refer to each types of the violated terms as X -type and Z -type. Since X -type term and Z -type term are lattice inversions of each other, it suffices to analyze just one of the two types. Not every configuration of defects is physically allowed. Indeed, a configuration of defects can be realized as an excited state if and only if it can be created by applying a tensor product P of Pauli matrices (Pauli operators) to the ground state. We call such defect configurations to be *valid*.

The most general energy eigenstate is a superposition of valid configurations with the same number of defects. Note that the configuration of defects does not uniquely determine the state. For example, the ground state has no defect, but is degenerate. Given the defect configuration, this is in fact the only residual degeneracy. We shall call this as the *topological degeneracy*.

Below, we will make use of the following fact¹⁶. Under periodic boundary conditions there is an infinite family of system sizes such that any defect configuration with

an even number of X -type defects and an even number of Z -type defects is valid, and the topological degeneracy is 4. This is the family that we shall study in this paper. Concretely, it includes the lattices of linear size $L = 2^n + 1$ for any integer $n \geq 1$. The situation is very similar to the toric code Hamiltonian; one cannot have an isolated anyon in a closed finite system, but any even number of anyons is allowed.

A. No-strings rule

The most important property, called *no-strings rule*¹⁷, is a formalization of the fact that isolated defects are immobile. More precisely, suppose $|\psi\rangle$ is a state with a defect at site i and no other defects within distance d from i . If $|\psi'\rangle$ is another state with a defect at $i' \neq i$ and no other defect within distance d from i , then the no-strings rule asserts that $\langle\psi'|O|\psi\rangle = 0$ for any operator O supported on a ball of diameter d . Note that the no-strings rule is a nontrivial condition, in that i' can be even in the vicinity of i .

The cubic code satisfies the no-strings rule, as we explain below. Note that any operator can be written as a linear combination of tensor products of Pauli matrices (Pauli operators). This is possible simply because Pauli matrices generate the full operator algebra. This observation eases the problem; in order to argue that no operator can “transport” a defect, it suffices to show that no Pauli operators can transport the defect. If a Pauli operator O supported on a ball of diameter d is capable of “transporting” a defect within that ball, some product of O and its translations can transport the defect arbitrarily far. The calculation of Ref. 15 implies that, if a defect can be transported arbitrarily far, it can be equally created from the ground state by acting some Pauli operator in its vicinity; see Fig. 2. This is impossible since a local Pauli operator can only create an even number of defects. We arrive at a contradiction, so we conclude that any Pauli operator, and hence any operator, is incapable of transporting defects.

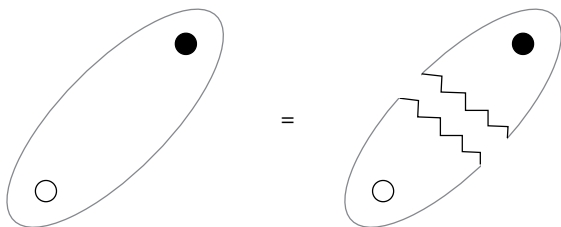


FIG. 2. No-strings rule. If some operator supported on a bounded region of the lattice can transport a defect, then there exists another operator that has exactly the same effect on the state and is supported on the vicinity of the initial and final defects. This is impossible since a single defect cannot be locally created or annihilated. Therefore, no such operator exists. See Ref. 15 for more detail.

The no-strings rule implies the existence of low-energy excited states which are locally gapped. To see this, consider an excited state $|\psi\rangle$ describing a configuration of defects that are separated from one another by a distance d . Let e be the total number of defects, so that the energy of the state is $2Je$. If another state of the same energy, denoted as $|\phi\rangle$, has a configuration of defects different from that of $|\psi\rangle$, there must be a defect at i that is present in $|\phi\rangle$ but not in $|\psi\rangle$. By the construction of $|\psi\rangle$, the no-strings rule implies that the matrix element $\langle\phi|O|\psi\rangle$ is zero for any operator O whose support is contained in a ball of diameter d . More generally, any state orthogonal to $|\psi\rangle$ with the same energy can be labeled by two components. One represents different defect configurations with the same energy, and the other represents topological degeneracy. The transition from $|\psi\rangle$ to any orthogonal and topologically degenerate state requires an operator whose support is comparable to the system size. By the no-strings rule, the transition from $|\psi\rangle$ to any other state with a different defect configuration requires an operator whose support is at least d . Other states have energy that is different from $2Je$ by at least $2J$. Therefore, the no-strings rule implies that a configuration of separated defects by distance d is locally gapped with diameter d and energy $2J$.

Remark that the existence of the locally gapped excited states is not due to any disorder; it follows from the no-strings rule, which is a nontrivial property of the cubic code model. A correct interpretation should be that defects at different locations represent distinct superselection sectors. A similar yet weaker version of this phenomenon is observed in Wen’s plaquette model¹⁸. There, a defect can be transported to next-nearest neighboring site by a local operator, but not to the nearest neighboring site. In other words, the superselection sector is changed under a unit translation, although it is not changed under two units of translations. In the cubic code model, there is no such finite periodicity, and consequently, there are infinitely many superselection sectors.

B. Typicality of locally gapped states

Now we show that almost all defect configurations are locally gapped. Without loss of generality, let us assume that m defects are all X -type, as the analysis for the Z -type defects are exactly the same. Any configuration is valid as long as m is even, and there is no restriction on the positions of the defects. The total number of states with a fixed energy will be given by the number of distinct configurations, multiplied by the topological degeneracy 4. This is a purely combinatorial problem in which one chooses m defect locations from a lattice of V sites such that every defect is separated from the others by distance d . Let us imagine that we are placing one defect at a time, eventually m defects in total. The first defect can be anywhere, as there are V possibilities. The second one can be anywhere except near the first one; there are

$V - v$ possibilities, where v is the volume of a ball of radius d . The third one can be placed anywhere among at least $V - 2v$ positions. Continuing in a similar way, we see that the number of these sparse defect configurations is at least $V(V - v)(V - 2v) \cdots (V - (m - 1)v)/m!$. The total number of all configurations of defects is $\binom{V}{m}$. Thus the fraction of the locally gapped states with diameter d is at least

$$\frac{V(V - v) \cdots (V - (m - 1)v)}{V(V - 1) \cdots (V - m + 1)} \geq \left(1 - \frac{mv}{V}\right)^m. \quad (3)$$

The fraction approaches 1 algebraically in the system size, provided that $d \sim L^{1-\epsilon}$ where $0 < \epsilon < 1$. We conclude that the majority of the defect configurations is locally gapped.

There are excited states which are not locally gapped, but they form a vanishing fraction of the low-energy subspace due to the previous analysis. These are the excited states with locally created clusters of defects. A local operator creates at least 4 defects from the ground state, and this cluster of 4 defects can hop to a nearby position by local operators. Thus, excited states with a topologically neutral (i.e., locally created) cluster of defects are not locally gapped, and may hybridize to become a momentum eigenstate upon perturbations.

IV. LOCALLY GAPPED STATES UNDER PERTURBATIONS

Locally gapped states are resilient to perturbations. Before we see why, let us recall a basic fact from perturbation theory to develop some intuitions. Consider a nondegenerate eigenstate $|\psi\rangle$ of some Hamiltonian H with an energy E . Under an infinitesimal perturbation Y , the eigenstate is perturbed as follows:

$$\delta|\psi\rangle = \sum_j \frac{\langle\psi_j|Y|\psi\rangle}{E - E_j} |\psi_j\rangle,$$

where $|\psi_j\rangle$ are the energy eigenstates of H with energy E_j . The correction term does not blow up if the state is either separated from the rest of the spectrum by a constant energy gap or the off-diagonal matrix elements are sufficiently small. These are exactly the two criteria that appear in the definition of locally gapped states. If the first criterion is satisfied, the state evolves under an adiabatic evolution, and we expect the wave function to change smoothly. If the second criterion is satisfied, the perturbation acts almost diagonally with respect to the original energy eigenstates, so we expect the effect of the perturbation to the eigenstates to be negligible. These observations suggest that if the energy profile of $|\psi\rangle$ is localized in the sense that there are a few positions around which energy density is nonzero, then the perturbed state also has a similar energy profile. This rough intuition will be made more precise in this Section.

A. Band-preserving perturbations

Consider a one-parameter family of Hamiltonians of the following form

$$H_s = H_0 + sY, \quad s \in [0, 1],$$

where H_0 is the original cubic code Hamiltonian (Eq. 2) and Y is some perturbation. We assume that the perturbation consists of local terms with bounded strength, where strength of the perturbation refers to the supremum of the norms of the local terms.

Let us invoke a powerful theorem which is due to Bravyi, Hastings, and Michalakis^{19,20}. Setting the energies of H_0 to be E_n ($n = 0, 1, 2, \dots$) with the ground state energy being $E_0 = 0$, the theorem says that the spectrum of H_s is contained in $[E_n(1 - cs), E_n(1 + cs)]$ up to a small correction that vanishes in the thermodynamic limit. Here, the number $c > 0$ is a constant that is independent of the system size, and the strength of the perturbation is assumed to be smaller than some constant. At finite energies, this implies that there are energy bands that are separated from each other by some constant, say Δ . This spectral gap condition allows us to unambiguously define projectors onto the linear span of states within one of the separated bands. We shall refer to such projectors as the projector onto the *band subspace*.

B. Evolution of the band subspace

The existence of the separated energy bands at finite energy implies that we can use a powerful machinery of quasi-adiabatic continuation^{21,22}. This allows us to systematically track the effect of the perturbation on these subspaces. It says that, whenever a band subspace around energy E_n is separated from the rest of the spectrum by a constant gap, the projector $P_s^{(n)}$ onto the band subspace obeys the following differential equation:

$$\partial_s P_s^{(n)} = i[K(s), P_s^{(n)}], \quad (4)$$

for some hermitian operator $K(s)$. Formally, the solution to this equation is the following:

$$P_s^{(n)} = U_s P_0^{(n)} U_s^\dagger \quad (5)$$

$$U_s = \mathcal{T} \exp \left(i \int_0^s K(u) du \right) \quad (6)$$

where \mathcal{T} is the “time”-ordering operation. The readers may wonder about the definition of $K(s)$ at this point. While its rather esoteric definition shall be given in the Appendix, its essential property can be summarized succinctly as follows.

The important point is that the unitary U_s preserves locality. Namely, for any operator O supported on a region M , we have the following decomposition:

$$U_s^\dagger O U_s = \sum_r O'_r, \quad \|O'_r\| \leq f(r), \quad (7)$$

where O'_r is some operator supported on a distance- r neighborhood of M and $f(r)$ is some function which decays rapidly in r .

Note that U_s governs the evolution of any energy band which is separated from the rest of the spectrum by a constant gap. As we have already discussed, the energy bands of the perturbed cubic code at finite energy satisfies this condition, provided that the strength of the perturbation is sufficiently small.

C. Approximately locally gapped states

By making use of these results, we can explicitly construct a large set of (approximately) locally gapped states. Such states are defined as follows:

$$|\psi_s\rangle = U_s |\psi_0\rangle, \quad (8)$$

where $|\psi_0\rangle$ is any state describing a sparse defect configuration of H_0 , the *unperturbed* cubic code Hamiltonian. Let us assume that (i) the perturbation strength is sufficiently small so that $|\psi_0\rangle$ belongs to some separated band, say $P_s^{(m)}$, and (ii) $|\psi_0\rangle$ is locally gapped with a diameter d .

These states are approximately locally gapped in the following sense. For the states that are not in the band, there is an energy gap Δ . For the states in the band, we need to consider the transition amplitude. Suppose there is an operator O supported on a ball of diameter $d/2$ such that $\langle\phi_s|O|\psi_s\rangle = \delta$ where $\langle\phi_s|\psi_s\rangle = 0$. This is to say that $\langle\phi_0|U_s^\dagger O U_s |\psi_0\rangle = \delta$ where $|\phi_0\rangle = U_s^\dagger |\phi_s\rangle$ is a state that belongs to $P_0^{(m)}$. Using the decomposition in Eq. (7), the operator $U_s^\dagger O U_s$ is quasi-local and hence can be approximated by an operator O' supported on the ball of diameter d such that $\|U_s^\dagger O U_s - O'\|$ decays rapidly in d . Since $|\psi_0\rangle$ is locally gapped with radius d , $\langle\phi_0|O'|\psi_0\rangle = 0$. It follows that δ decays rapidly with d . This establishes that $|\psi_s\rangle$ is approximately locally gapped with a diameter of $d/2$, up to an “error” that decays rapidly with d .

V. PROPERTIES OF LOCALLY GAPPED STATES

In the previous section, we have identified the locally gapped states of the perturbed Hamiltonian (H_s). In this section, we study their static and dynamical properties; we show that they have spatially localized energy profile, obey an area law of entanglement entropy, behave almost like energy eigenstates of H_s , and admit an extensive number of approximate local integrals of motion. Since the ratio of the number of localized states relative to the dimension of the low-energy subspace is very close to 1 from our counting argument, we conclude that a typical state from the low-energy subspace of arbitrarily perturbed Hamiltonian consists of “frozen” defects.

A. Static properties

Here, we show that the spatial energy profile of $|\psi_s\rangle$ is localized. We actually make a stronger claim: that the local reduced density matrix of $|\psi_s\rangle$ at a region M is close to that of the ground state of H_s if M is far from the locations of the defects for $|\psi_0\rangle$.

In order to see this, recall that the local reduced density matrix is completely determined by the expectation values of local observables.

$$\langle\psi_s|O|\psi_s\rangle = \langle\psi_0|U_s^\dagger O U_s |\psi_0\rangle.$$

Since U_s preserves locality, the “dressed” operator $U_s^\dagger O U_s$ can be well-approximated by some local operator, say O' . If this local approximation acts away from the defects, its expectation value with respect to $|\psi_0\rangle$ reproduces the ground state expectation value of the unperturbed system. If the ground states of the unperturbed and perturbed Hamiltonians are $|\Omega_0\rangle$ and $|\Omega_s\rangle$, respectively, the preceding discussion implies that $\langle\psi_s|O|\psi_s\rangle \simeq \langle\Omega_0|O'|\Omega_0\rangle = \langle\Omega_s|U_s O' U_s^\dagger |\Omega_s\rangle \simeq \langle\Omega_s|O|\Omega_s\rangle$. Therefore, any observable acting far from the defects cannot distinguish $|\psi_s\rangle$ from $|\Omega_s\rangle$. The error term decays rapidly in the distance from O to the defect’s location. (Recall that a rapid decaying function scales like $x^\alpha e^{-cx^\beta}$ for some constant $c, \beta > 0$ and α for large x .) In comparison, it is worth noting that the difference between the local reduced density matrices of a momentum eigenstate and that of the ground state can only be algebraically small ($x^{-\gamma}$ for some $\gamma > 0$) in the system size.

We note in passing that the state $|\psi_s\rangle$ obeys an area law of entanglement entropy. Namely, the entanglement entropy $S(M, |\psi_s\rangle)$ for a ball-like region M scales like the area of the boundary of M . This is not too surprising in light of a recent result²³; in fact, it has nothing to do with our construction of states with a localized energy profile. This is rather a consequence of the locality-preserving property of U_s and the smallness of the low energy band subspace. Consider first the unperturbed Hamiltonian H_0 . Every state $|\psi_0\rangle$ with a definite configuration of m defects, that is not necessarily sparse, has the same entanglement entropy for any region as the ground state, since it is related to the ground state by $|\psi_0\rangle = P |\Omega_0\rangle$ for some tensor product unitary operator P . Moreover, since the ground state is defined by a commuting set of Pauli operators, the nonzero Schmidt coefficients with respect to the region M are all the same. There are precisely 2^{S_0} nonzero Schmidt coefficients, where $S_0 = S(M, |\psi_0\rangle)/\log 2$. We consider the Schmidt decomposition of an arbitrary superposition $|\psi_0\rangle$ of the states within the band $P_0^{(m)}$ with respect to the region M and its complement. Since there are at most $\mathcal{O}(\text{Vol } M)^m$ states with different defect configurations in M , the Schmidt rank is at most $\mathcal{O}((\text{Vol } M)^m 2^{S_0})$. The topological degeneracy does not enter to this upper bound since topologically degenerate states have exactly the same reduced density matrix for a ball-like region M . The entanglement entropy is therefore at most

$S(M, |\psi_0\rangle) + m \log \text{Vol } M + O(1)$, which is an area law. The result of Ref. 23 states that for any state $|\psi\rangle$, the evolved state $U_s |\psi\rangle$ obeys an area law of entanglement entropy whenever $|\psi\rangle$ does. Since any state $|\psi_s\rangle$ in the band $P_s^{(m)}$ is equal to $U_s |\psi_0\rangle$ for some $|\psi_0\rangle$ in $P_0^{(m)}$, we have the area law for $|\psi_s\rangle$.

B. Dynamical properties

Now we are in a position to derive the main result, namely Eq. (1). The key insight is that the local Hamiltonian has small matrix elements between the locally gapped state $|\psi_s\rangle$ and the other states. To formalize this, let us fix a basis of the band subspace $P_s^{(m)}$ consisting of

$$|\tilde{1}\rangle = |\psi_s\rangle = U_s |\psi_0\rangle, \quad |\tilde{q}\rangle = U_s |q\rangle \quad (q > 1)$$

where $|q\rangle$ are energy eigenstate of H_0 with definite defect configurations. Here $|\tilde{1}\rangle$ is approximately locally gapped, but $|\tilde{q}\rangle$ may not be so. In the Appendix, we show that

$$\sum_{q \neq 1} |\langle q | H_s | \psi_s \rangle| \leq g(d) \quad (9)$$

where g is a rapidly decaying function. Since the matrix element is so small, the time evolution can be “simulated” by neglecting those small matrix elements. Indeed, let Q be a hermitian matrix acting within $P_s^{(s)}$, defined as

$$\langle \tilde{u} | Q | \tilde{v} \rangle = \begin{cases} 0 & \text{if } u = 1 \neq v \text{ or } u \neq 1 = v, \\ \langle \tilde{u} | H_s | \tilde{v} \rangle & \text{otherwise.} \end{cases} \quad (10)$$

Then, Eq. (9) implies that

$$\|P_s(Q - H_s)P_s\| \leq g(d), \quad (11)$$

where we used the fact that the operator norm is at most the sum of absolute values of the matrix elements. This in turn implies that

$$\begin{aligned} \|e^{-itA} - e^{-itB}\| &= \|1 - e^{itA}e^{-itB}\| \\ &= \left\| \int_0^t dw \partial_w (e^{iwA}e^{-iwB}) \right\| \\ &\leq \int_0^{|t|} dw \|e^{iwA}(A - B)e^{-iwB}\| \leq |t|g(d) \end{aligned}$$

where $A = P_s H_s P_s$ and $B = P_s Q P_s$, from which we conclude that

$$|\langle \psi_s | e^{-itH_s} | \psi_s \rangle| \geq 1 - |t|g(d).$$

This is Eq. (1).

The fact that these localized states remain almost invariant under time evolution is closely related to the existence of the local integrals of motion. The definition of local integrals of motion in our context is a bit

more relaxed than those that are discussed in the context of many-body localization. Since we are studying a finite-energy subspace, the integrals of motion of our interest may not commute with the Hamiltonian on the entire Hilbert space. In fact, since our system can be translation-invariant, we should not expect such commutation relation in general. However, there are local operators that commute with our Hamiltonian within the localized subspace, e.g., the subspace spanned by the locally gapped states of the perturbed Hamiltonian.

More concretely, let us define P_{loc} be the projector onto the localized subspace. This refers to the linear span of $U_s |q\rangle$, where $|q\rangle$ describes a state with defects which are separated by a distance d . Define

$$\mathcal{I}_j(s) = U_s h_j^{(0)} U_s^\dagger \quad (12)$$

where $h_j^{(0)}$ is the local term of the Hamiltonian H_0 . This operator is quasi-local due to the locality-preserving property, i.e., Eq. (7).

Now let us see whether this operator commutes with the Hamiltonian within the localized subspace.

$$\begin{aligned} \|P_{\text{loc}}[H_s, \mathcal{I}_j(s)]P_{\text{loc}}\| &\leq \sum_{a,b} \left| \langle \tilde{a} | [H_s, \mathcal{I}_j(s)] | \tilde{b} \rangle \right| \\ &= \sum_{a,b} \left| (\ell_a - \ell_b) \langle \tilde{a} | H_s | \tilde{b} \rangle \right| \\ &\leq \sum_{a \neq b} 2 \left| \langle \tilde{a} | H_s | \tilde{b} \rangle \right| \leq 2(\dim P_{\text{loc}})g(d) \\ &\leq 2L^{3m}g(d). \end{aligned} \quad (13)$$

where L is the linear system size. The first inequality is the triangle inequality, the second one follows from $\|\mathcal{I}_j(s)\| = 1$, and the third one is from Eq. (9). If we choose $d \sim L^{1-\epsilon}$ for a small $\epsilon \in (0, 1)$, then the upper-bound decays rapidly in the system size, and Eq. (3) reads

$$\frac{\dim P_{\text{loc}}}{\dim P_s^{(m)}} \geq 1 - L^{-\beta}.$$

for some $\beta > 0$. Thus, $\mathcal{I}_j(s)$ is almost an integral of motion for almost all states of the band. Indeed, the expectation value of $\mathcal{I}_j(s)$ with respect to locally gapped states is slowly varying. Let $|\psi\rangle$ be *any* state in the subspace P_{loc} .

$$\begin{aligned} \langle \psi | e^{itH_s} \mathcal{I}_j(s) e^{-itH_s} | \psi \rangle &= \langle \psi | \mathcal{I}_j(s) | \psi \rangle + \int_0^t dw \langle \psi | e^{iwH_s} i[H_s, \mathcal{I}_j(s)] e^{-iwH_s} | \psi \rangle \\ &= \langle \psi | \mathcal{I}_j(s) | \psi \rangle \pm \int_0^t dw \mathcal{O}(\|P_{\text{loc}}[H_s, \mathcal{I}_j(s)]P_{\text{loc}}\|) \\ &= \langle \psi | \mathcal{I}_j(s) | \psi \rangle \pm t\mathcal{O}(L^\alpha g(d)) \end{aligned}$$

The expectation value of $\mathcal{I}_j(s)$ stays at the initial value up to a time that grows rapidly in the system size.

VI. DISCUSSION

There is a considerable amount of work in the literature regarding the fate of the disorder-driven localized system in the presence of interaction; see e.g. review articles Ref. 24 and 25. Here we have taken a different route, and identified a mechanism for localization which is driven by a strong interaction. By exploiting the fact that the ground state of our model is topologically ordered, we rigorously showed that the localized energy profile of the majority of the low energy states remains unchanged under arbitrary perturbations.

We remark that, though our analysis was focused on one particular model, a similar phenomenon is also observed in other models such as Chamon's spin model^{26,27} and other "cubic code" models^{15,28,29}. In Chamon's model, there exists a string operator that is capable of moving topologically nontrivial excitation. However, there exists an isolated excitation such that no operator can move it without creating extra excitations. A sparse configuration of these isolated excitations is a locally gapped state, and our analysis carries over.

Our work strongly suggests that the prevalent dichotomous view on the role of disorder and interaction needs to be modified. That is, certain strong enough interactions can drive the system into an exotic topologically ordered phase, in which localization occurs at low energy.

An outstanding open problem is whether a locally gapped state can exist at finite energy density, and remain stable against perturbations. At such high energy, the gap between the bands would collapse upon perturbation, and the machinery of quasi-adiabatic continuation in its present form cannot be applied.

ACKNOWLEDGMENTS

We thank T. Senthil for bringing our attention to this problem, and also Soonwon Choi for insightful discussions. JH is supported by Pappalardo Fellowship in Physics at MIT. IK's research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

Appendix A: Generator of quasi-adiabatic continuation

The quasi-adiabatic continuation operator(U_s) obeys the following differential equation:

$$\frac{d}{ds}U_s = iK(s)U_s, \quad (\text{A1})$$

where $U_0 = I$. Here we explicitly write down the generator, namely $K(s)$. While there can be many different choices of the generator, the specific choice we make is

due to Bachmann *et al.*³⁰:

$$K(s) = \int_{-\infty}^{\infty} w_{\gamma}(t) \int_0^t du e^{iuH(s)} Y e^{-iuH(s)}. \quad (\text{A2})$$

Here $w_{\gamma}(t)$ is a normalized and rapidly decaying function whose Fourier transform is supported on $[-\gamma, \gamma]$. Such a function is known to exist. An example is given in Ref. 30, which decays faster than $\exp(-c\gamma t / \log^2(\gamma t))$ for some $c > 0$. If the energy gap between a band and the rest of the spectrum is bounded from below by γ , $K(s)$ generates the unitary flow of the band subspace along the path. The rapid decay of $w_{\gamma}(t)$ implies that U_s preserves locality. In particular, it is known that U_s maps a local operator to a quasi-local operator, i.e., an operator that can be well-approximated by a local operator.^{30,31}

Appendix B: Proof of Eq. (9)

Consider the off-diagonal element of the Hamiltonian between two orthogonal states within the band subspace.

$$\langle \tilde{q} | H_s | \tilde{1} \rangle = \sum_i \langle q | U_s^\dagger h_i U_s | 1 \rangle$$

The term $U_s^\dagger h_i U_s$ is quasi-local, and thus can be decomposed as $\sum_{r \geq 0} h'_{i,r}$, where $h'_{i,r}$ is supported on a ball of diameter r centered at i such that $\|h'_{i,r}\| \leq f(r)$ for a rapidly decaying function f . We need to consider two cases.

Firstly, if $|q\rangle$ and $|1\rangle$ have the same defect configuration, i.e., they are topologically degenerate, then $h_{i,r}$ must have support across the system; $r \sim L$. The topological degeneracy given a defect configuration is 4 for the cubic code under the aforementioned boundary conditions. Thus, the contribution from the topological degeneracy is at most $4f(L)$.

Secondly, suppose $|q\rangle$ and $|1\rangle$ have different defect configurations. For each i , let $j(i, q)$ be the furthest location from i at which the defect is present, *only* in one of the two states. Each off-diagonal terms, i.e., $\langle q | h'_{i,r} | 1 \rangle$, vanish unless $h'_{i,r}$ creates or annihilates a defect at $j = j(i, q)$. In order for this to happen, r has to be larger or equal to the distance between i and j . Therefore, if $\text{dist}(i, j)$ is large, the off-diagonal terms decay rapidly. If $\text{dist}(i, j)$ is small, a better bound can be obtained as follows. Since $|1\rangle$ is locally gapped with a distance d , $\langle q | h'_{i,r} | 1 \rangle$ vanishes for $r < d$. These two cases cover all the possibilities, and we conclude that the nonzero contributions come from $h'_{i,r}$ with $r \geq R(i, q) = \max(\text{dist}(i, j), d)$:

$$|\langle q | U_s^\dagger h_i U_s | 1 \rangle| \leq \sum_{r \geq R(i, q)} f(r) =: f_1(R(i, q)).$$

We now sum the right-hand side over all the states q and terms h_i , to obtain Eq. (9).

Suppose m is the number of defects in $|1\rangle$. Let us upper-bound the number of tuples (i, q) such that

$\text{dist}(i, j(i, q)) \leq x$ for a given x . The defects of $|q\rangle$, as well as the site i , has to be contained in a $2x$ -neighborhood of the defects of $|1\rangle$; otherwise the distance between i and j is larger than x . The number of such configurations is

at most $((2x)^3)^m \cdot m(2x)^3 \sim x^{3m+3}$. Then,

$$\begin{aligned} & \sum_{i,q} f_1(R(i, q)) \\ &= \sum_{i,q:\text{dist}(i,j)\leq d} f_1(d) + \sum_{i,q:\text{dist}(i,j)=x>d} f_1(x) \\ &\leq O(1) \sum_{x\geq d} x^{3m+3} f_1(x), \end{aligned}$$

where in the last line we made use of the fact that the topological degeneracy is finite. Accounting for the topological degeneracy, we get

$$\sum_{q \neq 1} |\langle q | H_s | 1 \rangle| \leq 4f(L) + O(1) \sum_{x \geq d} x^{3m+3} f_1(x). \quad (\text{B1})$$

Both f and f_1 are rapidly decaying. Eq. (9) is proved.

-
- ¹ J. M. Deutsch, *Phys. Rev. A* **43**, 2046 (1991).
² M. Srednicki, *Phys. Rev. E* **50**, 888 (1994), [cond-mat/9403051](#).
³ P. W. Anderson, *Phys. Rev.* **109**, 1492 (1958).
⁴ D. M. Basko, I. L. Aleiner, and B. L. Altshuler, *Annals of Physics* **321**, 1126 (2006), [cond-mat/0506617](#).
⁵ A. Pal and D. A. Huse, *Phys. Rev. B* **82**, 174411 (2010), [1010.1992](#).
⁶ M. Serbyn, Z. Papić, and D. A. Abanin, *Phys. Rev. Lett.* **111**, 127201 (2013).
⁷ D. A. Huse and V. Oganesyan, *Phys. Rev. B* **90**, 174202 (2013), [1305.4915](#).
⁸ T. Grover and M. P. A. Fisher, *J. Stat. Mech.* **P10010** (2014).
⁹ M. Schiulaz and M. Mller, *AIP Conf. Proc.* **1610**, 11 (2013), [1309.1082](#).
¹⁰ W. D. Roeck and F. Huveneers, *Comm. Math. Phys.* **332**, 1017 (2014), [1308.6263](#).
¹¹ J. M. Hickey, S. Genway, and J. P. Garrahan, (2014), [1405.5780](#).
¹² M. Schiulaz, A. Silva, and M. Mller, (2014), [1410.4690](#).
¹³ N. Y. Yao, C. R. Laumann, J. I. Cirac, M. D. Lukin, and J. E. Moore, (2014), [1410.7407](#).
¹⁴ Z. Papic, E. M. Stoudenmire, and D. A. Abanin, (2015), [1501.00477](#).
¹⁵ J. Haah, *Phys. Rev. A* **83**, 042330 (2011), [1101.1962](#).
¹⁶ J. Haah, *Commun. Math. Phys.* **324**, 351 (2013), [1204.1063](#).
¹⁷ S. Bravyi and J. Haah, *Phys. Rev. Lett.* **107**, 150504 (2011), [1105.4159](#).
¹⁸ X.-G. Wen, *Phys. Rev. Lett.* **90**, 016803 (2003), [quant-ph/0205004](#).
¹⁹ S. Bravyi, M. Hastings, and S. Michalakis, *J. Math. Phys.* **51**, 093512 (2010), [1001.0344](#).
²⁰ S. Bravyi and M. B. Hastings, *Comm. Math. Phys.* **307**, 609 (2011), [1001.4363](#).
²¹ M. B. Hastings and X.-G. Wen, *Phys. Rev. B* **72**, 045141 (2005), [cond-mat/0503554](#).
²² T. J. Osborne, *Phys. Rev. A* **75**, 032321 (2007).
²³ K. Van Acoleyen, M. Mariën, and F. Verstraete, *Phys. Rev. Lett.* **111**, 170501 (2013).
²⁴ R. Nandkishore and D. A. Huse, *Annual Review of Condensed Matter Physics* **6**, 15 (2015).
²⁵ C. Gogolin and J. Eisert, (2015), [1503.07538](#).
²⁶ C. Chamon, *Phys. Rev. Lett.* **94**, 040402 (2005), [cond-mat/0404182](#).
²⁷ S. Bravyi, B. Leemhuis, and B. M. Terhal, *Annals of Physics* **326**, 839 (2011), [1006.4871](#).
²⁸ I. H. Kim, (2012), [1202.0052](#).
²⁹ B. Yoshida, *Phys. Rev. B* **88**, 125122 (2013).
³⁰ S. Bachmann, S. Michalakis, B. Nachtergaele, and R. Sims, *Commun. Math. Phys.* **309**, 835 (2012), [1102.0842](#).
³¹ M. B. Hastings, *Phys. Rev. B* **69**, 104431 (2004), [cond-mat/0305505](#).